Metricas

Métricas de performance

Notaciones

- M_i es la i-ésima observación de clorofila **medida** in-situ (en μg/L).
- E_i es la i-ésima **estimación** del valor de clorofila hecha (a partir de un *único* modelo no especificado) (también en μg/L).
- $E_{i|m}$ es la i-ésima estimación del valor de clorofila hecha a partir del modelo m (también en μg/L).
- K es el conjunto de todos los modelos posibles (o sea, $m \in K$). No usé la letra M porque ya estaba tomada :(
- $mean(M)$, $median(M)$ y $sd(M)$ son el promedio, la mediana y el desvío estándar, respectivamente, del conjunto de valores M_i. Análogamente para los $E_{i|m}$. También puede figurar como $median(M_i)$.
- $sign(X)$ es la función signo: 1 si X > 0 y -1 si X < 0

MWR

Es el porcentaje de las observaciones en las que "gana" cada modelo, en donde ganar significa hacer la predicción con menor valor absoluto. La definición matemática sería:

$$MRW(\%) = 100 \times \frac{1}{N} \sum_{i..N} p(i,m)$$

En donde:

$$p(i,m) = \begin{cases} 1 & \text{si } |E_{i|m} - M_i| = min\{|E_{i|m} - M_i|, \forall m \in K\} \\ 0 & \text{en caso contrario} \end{cases}$$

Explicación encontrada en Seegers et al. 2018:

2.2.2 Decision metrics

Decision metrics enable additional comparison and selection of algorithms. Decision metrics date back to the 18th century mathematician Condorcet and are often described as "voting" methods [32]. One immediate practical approach is the pair-wise comparison based on Condorcet [33]. Pair-wise comparisons operate sequentially on each observation: (1) for a given observation, the model-observation differences are calculated for every model under consideration; (2) the model with the minimal difference is designated the winner for that given observation; (3) the number of wins per model are tabulated for all observations; and, (4) the model with the most wins is designated the best performing model. Unlike many other error metrics, the pair-wise comparison directly considers model failures – when model *A* provides a valid retrieval for a given observation but model *B* does not, only model *A* remains in the pool of potential winners for that observation. This metric will penalize a model that fails frequently, but performs well when it works. In this study, we adopted the pair-wise comparison of algorithm residuals (= model – observation), with the lowest residual designated as the winner. Results of this analysis were reported in terms of *percent wins*.

AEmean

"Mean Absolute Error", tomada de Zhang et al. 2015. Es decir, el error absoluto promedio:

$$AE_{mean} = \frac{1}{N} \sum_{i..N} |E_i - M_i|$$

REmean

"Mean Relative Error", tomada de Zhang et al. 2015. El error relativo (a la medición in-situ) promedio:

$$RE_{mean} = \frac{1}{N} \sum_{i..N} |E_i - M_i|/M_i$$

DMC

"Standard Deviaton from the Mean", tomada de Zhang et al. 2015... Es decir, qué tan lejos está el promedio de las predicciones, en relación al promedio de las medidas in-situ, normalizado por ese último:

$$DMC(\%) = 100 \times (mean(E) - mean(M))/mean(M)$$

DSD

"Standard Deviation of the Standard Deviation", tomada de Zhang et al. 2015. Igual que el anterior, pero con el desvío estándar.

$$DSD(\%) = 100 \times (sd(E) - sd(M))/sd(M)$$

Epsilon y Beta

Propuestas por Morley et al. 2018 (ver mail de Nima 18/11/2020). Son la "Median Symmetric Accuracy" (MdSA) y el "Symmetric Signed Percentage Bias".

$$\epsilon(\%) = 100 \times (10^{Y} - 1) \quad \text{donde} \quad Y = median(|log_{10}(E_i/M_i)|)$$

$$\beta(\%) = 100 \times sign(Z) \times (10^{Z} - 1) \quad \text{donde} \quad Z = median(log_{10}(E_i/M_i))$$

Nota: $log_{10}(E_i/M_i) = log_{10}E_i - log_{10}M_i$, por lo que estas fórmulas se parecen a las de MAE y Bias encontradas en Pahlevan et al 2019.

RMSE y RMSLE

RMSE: Root Mean Squared Error

RMSLE: Root Mean Squared Log Error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i..N}(E_i - M_i)^2}$$

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i..N}(log_{10}E_i - log_{10}M_i)^2}$$

MAPE

Median Absolute Percentage Error. Es muy parecido al REmean, pero usa la mediana en lugar del promedio del error relativo:

$$MAPE(\%) = 100 \times median\left(|E_i - M_i|/M_i\right)$$

Bias

Traducido como sesgo: "log transformed residuals". En su "intención" es parecido al epsilon (ϵ) pero usa el promedio en lugar de la mediana:

$$Bias = 10^{Z} \quad \text{donde} \quad Z = \frac{1}{N}\sum_{i..N}(log_{10}E_i - log_{10}M_i)$$

MAE

Mean Absolute Error comuted on log-scale. En su "intención" es parecido al beta (β) pero usa el promedio en lugar de la mediana:

$$MAE = 10^{Y} \quad \text{donde} \quad Y = \frac{1}{N}\sum_{i..N}|log_{10}E_i - log_{10}M_i|$$

MWRp

Model Win Rate **basado en las métricas de desempeño** (o *performance*). Análogo a MWR, pero esta vez hace un ranking entre modelos basado en un conjunto de métricas de desempeño. Es decir, un MWRp = 50% indica que el modelo m es el mejor para el 50% de las métricas evaluadas.

Se puede expresar en una ecuación para el MWRp de un modelo m determinado, siendo $\omega(E_m)$ el valor de la métrica ω para el conjunto de estimaciones de Clorofila E_m generadas por el modelo:

$$MRW(\%) = 100 \times \frac{1}{L} \sum_{i..L} D(\omega)$$

En donde L es el total de métricas incluidas en el cálculo y:

$$D(\omega) = \begin{cases} 1 & \text{si } \omega(E_m) \text{ es el mejor desempeño en } \{\omega(E_m), \forall m \in K\} \\ 0 & \text{en caso contrario} \end{cases}$$

Las performances utilizadas en la evaluación de MWRp son AEmean, REmean, RMSE, RMSLE, Beta, Epsilon, MWR, DMC y DSD
